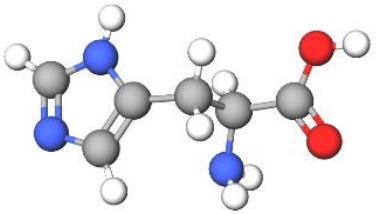
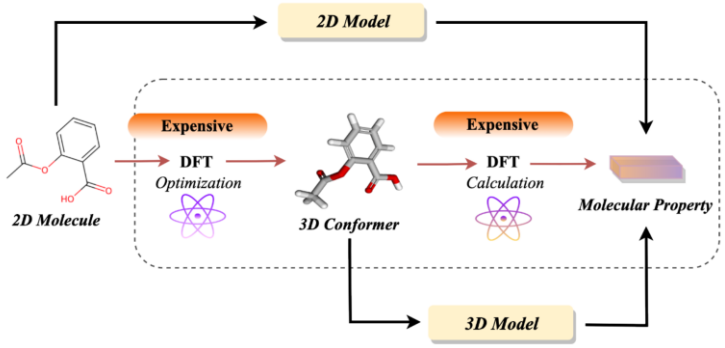
Denoising-based 3D Molecular Pre-training

12 June 2024

Preliminary



- 3D molecular structure: a point cloud of atomic nuclei in $\mathbb{R}^3$
- Molecular property prediction tasks based on 3D molecular structures
  - QM9: geometric, energetic, electronic and thermodynamic properties for 134k stable small organic molecules made up of CHONF
  - MD17: molecular dynamics trajectories of 8 small organic molecules
  - OC20: the relaxed energy from the initial structure (IS2RE)
  - DES15K: the interaction energy of molecular pairs

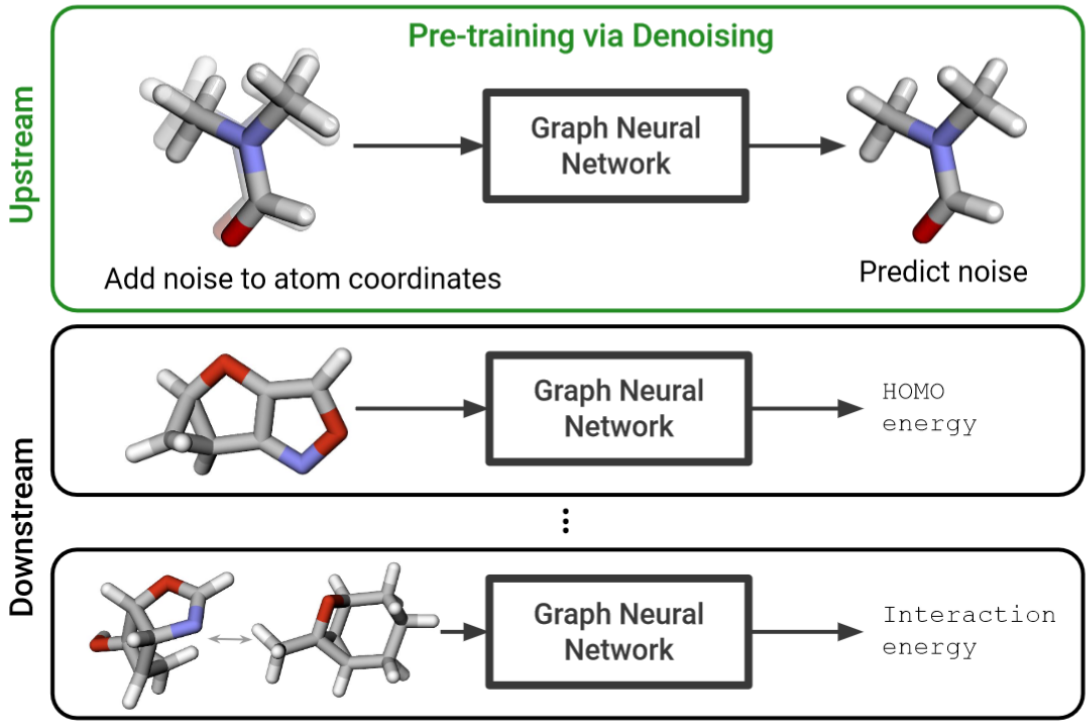


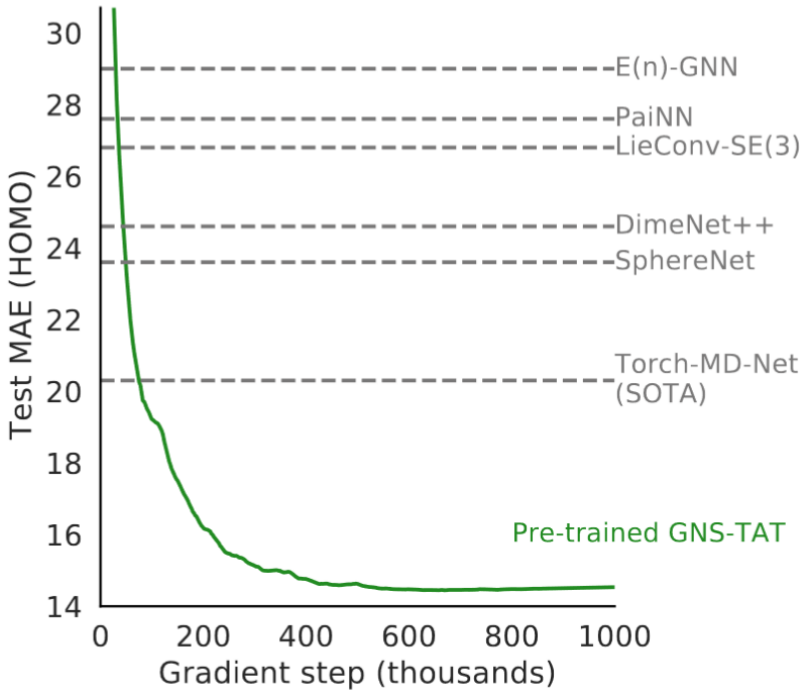
- 3D molecular pre-training
  - Dataset: PCQM4Mv2 (3.4 million organic molecules, with one equilibrium conformation and one label calculated by DFT)
  - Pre-training task: denoising

Related Work

- [1] Sheheryar Zaidi, Michael Schaarschmidt, James Martens, Hyunjik Kim, Yee Whye Teh, Alvaro Sanchez-Gonzalez, Peter Battaglia, Razvan Pascanu, Jonathan Godwin. **"Pre-Training via Denoising for Molecular Property Prediction."** In *ICLR*, 2023
- [2] Gengmo Zhou, Zhifeng Gao, Qiankun Ding, Hang Zheng, Hongteng Xu, Zhewei Wei, Linfeng Zhang, Guolin Ke. **"Uni-Mol: A Universal 3D Molecular Representation Learning Framework."** In *ICLR*, 2023
- [3] Shengchao Liu, Hongyu Guo, Jian Tang. **"Molecular Geometry Pretraining with SE(3)-Invariant Denoising Distance Matching."** In *ICLR*, 2023
- [4] Shengjie Luo, Tianlang Chen, Yixian Xu, Shuxin Zheng, Tie-Yan Liu, Liwei Wang, Di He. **"One Transformer Can Understand Both 2D & 3D Molecular Data."** In *ICLR*, 2023
- [5] Rui Jiao, Jiaqi Han, Wenbing Huang, Yu Rong, Yang Liu. **"Energy-Motivated Equivariant Pretraining for 3D Molecular Graphs."** In *AAAI*. 2023
- [6] Hyeonsu Kim, Jeheon Woo, Seonghwan Kim, Seokhyun Moon, Jun Hyeong Kim, Woo Youn Kim. **"GeoTMI:Predicting Quantum Chemical Property with Easy-to-Obtain Geometry via Positional Denoising."** In *NeurIPS*. 2023
- [7] Shikun Feng, Yuyan Ni, Yanyan Lan, Zhi-Ming Ma, Wei-Ying Ma. **"Fractional Denoising for 3D Molecular Pre-Training."** In *ICML*. 2023
- [8] Yuyan Ni, Shikun Feng, Wei-Ying Ma, Zhi-Ming Ma, Yanyan Lan. **"Sliced Denoising: A Physics-Informed Molecular Pre-Training Method."** In *ICLR*, 2024

Coordinate Denoising, ICLR 2023





- Method: $\tilde{S} = \{(a_1, \tilde{\mathbf{p}}_1), \ldots, (a_{|S|}, \tilde{\mathbf{p}}_{|S|})\}$, where $\tilde{\mathbf{p}}_i = \mathbf{p}_i + \sigma\boldsymbol{\epsilon}_i$ and $\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, I_3)$,

$$\mathbb{E}_{p(\tilde{S},S)}\left[\left\|\mathrm{GNN}_\theta(\tilde{S}) - (\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_{|S|})\right\|^2\right].$$

- Denoising as learning a force field.

Coordinate Denoising, ICLR 2023

- Denoising as learning a force field.
  - It is not feasible to learn the molecular force field directly, because it is either unknown or expensive to evaluate.
  - Alternative: approximate the data-generating force field with one that can be cheaply evaluated.
  - Prove that the denoising objective is equivalent to learning the molecular force field:
    - Molecular structure: $\mathbf{x} \in \mathbb{R}^{3N}$
    - The structure follows the Boltzmann distribution: $p_{\text{physical}}(\mathbf{x}) \propto \exp(-E(\mathbf{x}))$
    - Force field: $\nabla_{\mathbf{x}} \log p_{\text{physical}}(\mathbf{x}) = -\nabla_{\mathbf{x}} E(\mathbf{x})$
    - Approximate p_{physical} with a mixture of Gaussians centered at the known equilibrium structures

$$p_{\text{physical}}(\tilde{\mathbf{x}}) \approx q_\sigma(\tilde{\mathbf{x}}) := \frac{1}{n} \sum_{i=1}^{n} q_\sigma(\tilde{\mathbf{x}} \mid \mathbf{x}_i)$$

where $q_\sigma(\tilde{\mathbf{x}} \mid \mathbf{x}_i) = \mathcal{N}(\tilde{\mathbf{x}}; \mathbf{x}_i, \sigma^2 I_{3N})$

Coordinate Denoising, ICLR 2023

- Denoising as learning a force field. (Cont.)
  - Learning the force field now yields a score-matching objective:

$$\mathbb{E}_{q_\sigma(\tilde{\mathbf{x}})}\left[\left\|\mathrm{GNN}_\theta(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} \log q_\sigma(\tilde{\mathbf{x}})\right\|^2\right] \tag{1}$$

  - According to reference [1], minimizing the following two objectives is equivalent:

$$J_1(\theta) = \mathbb{E}_{q_\sigma(\tilde{\mathbf{x}})}\left[\left\|\mathrm{GNN}_\theta(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} \log q_\sigma(\tilde{\mathbf{x}})\right\|^2\right]$$

$$J_2(\theta) = \mathbb{E}_{q_\sigma(\tilde{\mathbf{x}},\mathbf{x})}\left[\left\|\mathrm{GNN}_\theta(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} \log q_\sigma(\tilde{\mathbf{x}} \mid \mathbf{x})\right\|^2\right]$$

  - Thus, the objective in Eq. (1) is equivalent to:

$$\mathbb{E}_{q_\sigma(\tilde{\mathbf{x}},\mathbf{x})}\left[\left\|\mathrm{GNN}_\theta(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} \log q_\sigma(\tilde{\mathbf{x}} \mid \mathbf{x})\right\|^2\right] = \mathbb{E}_{q_\sigma(\tilde{\mathbf{x}},\mathbf{x})}\left[\left\|\mathrm{GNN}_\theta(\tilde{\mathbf{x}}) - \frac{\mathbf{x}-\tilde{\mathbf{x}}}{\sigma^2}\right\|^2\right]$$

[1] Pascal Vincent. **“A Connection Between Score Matching and Denoising Autoencoders.”** *Neural Computation* July 2011

Coordinate Denoising, ICLR 2023

Target	Unit	SchNet	E(n)-GNN	DimeNet++	SphereNet	PaiNN	TorchMD-NET	GNS + NN	GNS-TAT + NN	Pre-trained GNS-TAT + NN
μ	D	0.033	0.029	0.030	0.027	0.012	**0.011**	0.025	0.021	0.016
α	a_0^3	0.235	0.071	0.043	0.047	0.045	0.059	0.052	0.047	**0.040**
ϵ_{HOMO}	meV	41.0	29.0	24.6	23.6	27.6	20.3	20.4	17.3	**14.9**
ϵ_{LUMO}	meV	34.0	25.0	19.5	18.9	20.4	18.6	17.5	17.1	**14.7**
$\Delta\epsilon$	meV	63.0	48.0	32.6	32.3	45.7	36.1	28.6	25.7	**22.0**
$\langle R^2 \rangle$	a_0^2	0.07	0.11	0.33	0.29	0.07	**0.033**	0.70	0.65	0.44
ZPVE	meV	1.700	1.550	1.210	1.120	1.280	1.840	1.160	1.080	**1.018**
U_0	meV	14.00	11.00	6.32	6.26	5.85	6.15	7.30	6.39	**5.76**
U	meV	19.00	12.00	6.28	7.33	5.83	6.38	7.57	6.39	**5.76**
H	meV	14.00	12.00	6.53	6.40	5.98	6.16	7.43	6.42	**5.79**
G	meV	14.00	12.00	7.56	8.00	7.35	7.62	8.30	7.41	**6.90**
c_{v}	$\frac{\text{cal}}{\text{mol K}}$	0.033	0.031	0.023	0.022	0.024	0.026	0.025	0.022	**0.020**

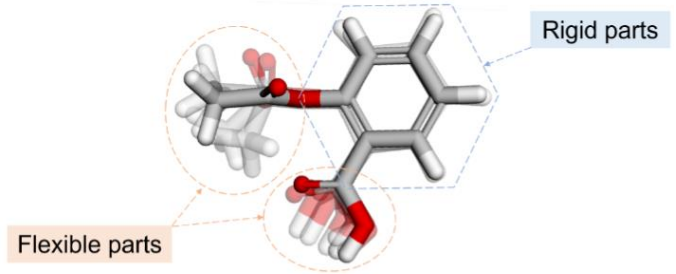
Frad (Fractional Denoising), ICML 2023

$$
\begin{aligned}
& E_{p(\tilde{x})}\|GNN_\theta(\tilde{x}) - (-\nabla_{\tilde{x}} E(\tilde{x}))\|^2 \quad \text{(a)} \\
=& E_{p(\tilde{x})}\|GNN_\theta(\tilde{x}) - \nabla_{\tilde{x}} \log p(\tilde{x})\|^2 \quad \text{(b)} \\
=& E_{p(\tilde{x}|x_i)p(x_i)}\|GNN_\theta(\tilde{x}) - \nabla_{\tilde{x}} \log p(\tilde{x}|x_i)\|^2 + T \quad \text{(c)} \\
=& E_{p(\tilde{x}|x_i)p(x_i)}\|GNN_\theta(\tilde{x}) - \frac{x_i - \tilde{x}}{\tau^2}\|^2 + T \quad \text{(d)},
\end{aligned}
$$

$$
\min_\theta E_{p(\tilde{x}|x_i)p(x_i)}\|GNN_\theta(\tilde{x}) - (\tilde{x} - x_i)\|^2
$$

Two challenges prevent the current coordinate denoising methods from learning an accurate force field:

- **Low Sampling Coverage.**
- **Isotropic Force Field.** $\Sigma = \tau^2 I_{3N}$

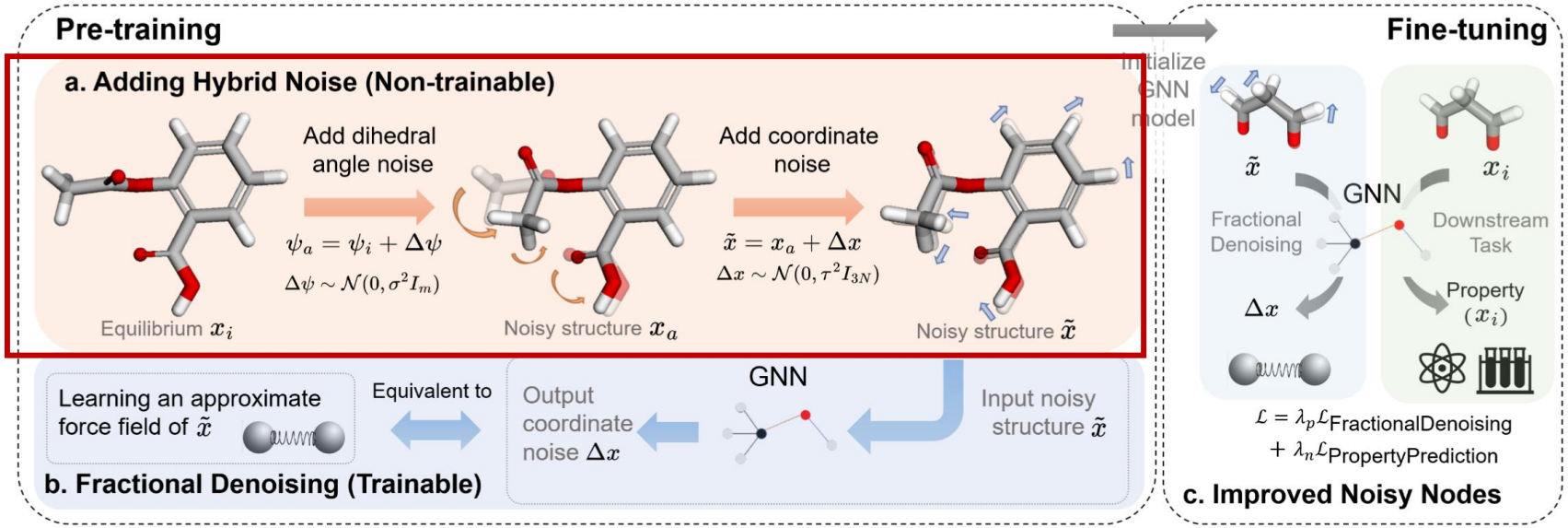


anisotropy of molecular structures

	$\tau = 0.004$	$\tau = 0.04$	$\tau = 0.4$	$\sigma = 1, \tau = 0.04$	$\sigma = 2, \tau = 0.04$	$\sigma = 20, \tau = 0.04$
Perturbation Scale	0.00319	0.0319	0.319	0.0635	0.0952	0.6499
Average Performance(meV)	14.06	12.86	15.26	12.06	11.94	12.30

Frad (Fractional Denoising), ICML 2023

Hybrid Noise



Difficulties to Learn the Force Field of Hybrid Noise

Table 1. The conditional distributions of the conformation perturbed by various noise type given the clean (equilibrium) conformation x_i and the corresponding score functions.

Noise Type	Conformation Distribution $p(\cdot\|x_i)$	Score Function $\nabla \log p(\cdot\|x_i)$
Coordinate	$x_{cd} \sim \mathcal{N}(x_i, \tau^2 I_{3N})$	$\frac{1}{\tau^2}(x_i - x_{cd})$
Dihedral Angle	$x_a \sim \mathcal{N}(x_i, \Sigma_\sigma)$	$\Sigma_\sigma^{-1}(x_i - x_a)$
Hybrid	$\tilde{x} \sim \mathcal{N}(x_i, \Sigma_{\sigma,\tau})$	$\Sigma_{\sigma,\tau}^{-1}(x_i - \tilde{x})$

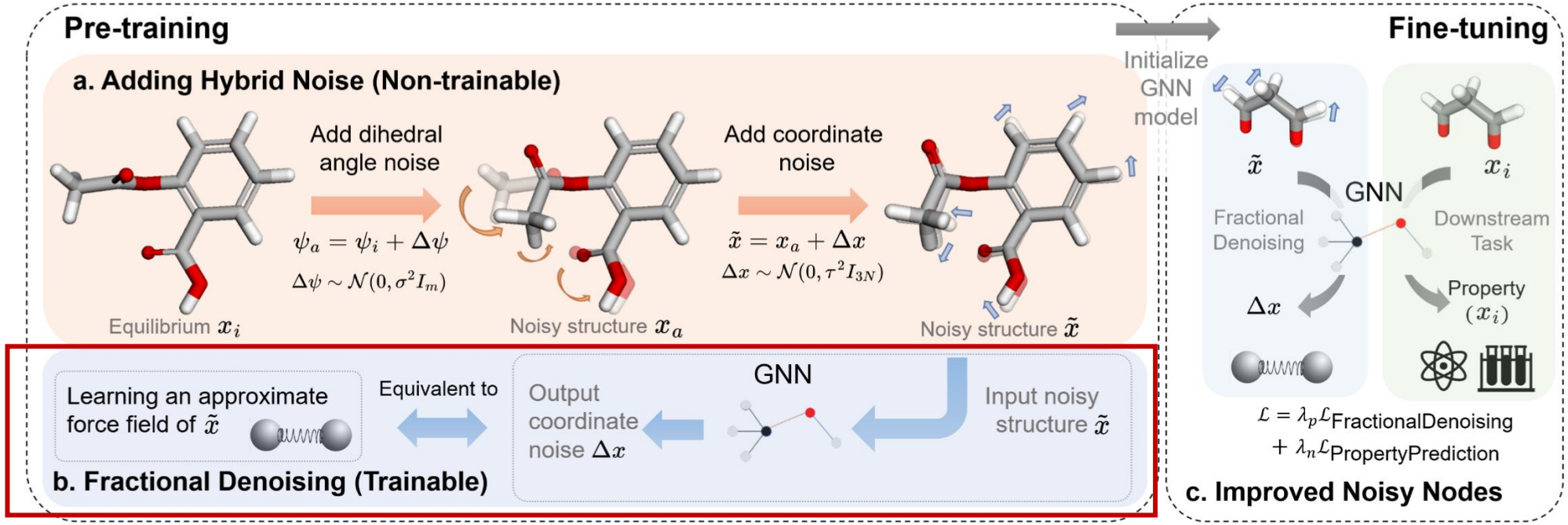
$$\|\Delta x - C\Delta\psi\|_2^2 \le \sum_{j=1}^{m} D_j \mathcal{E}(\Delta\psi_j)$$

Where C is a $3N \times m$ matrix that is dependent on the input conformation

$$\Sigma = \sigma^2 CC^T \qquad \Sigma_{\sigma,\tau} = \tau^2 I_{3N} + \sigma^2 CC^T$$

Neither denoising dihedral angle noise nor denoising hybrid noise is equivalent to learning the force field, because the coefficients $\Sigma_\sigma, \Sigma_{\sigma,\tau}$ rely on input conformation and cannot be absorbed into GNN_θ.

Frad (Fractional Denoising), ICML 2023

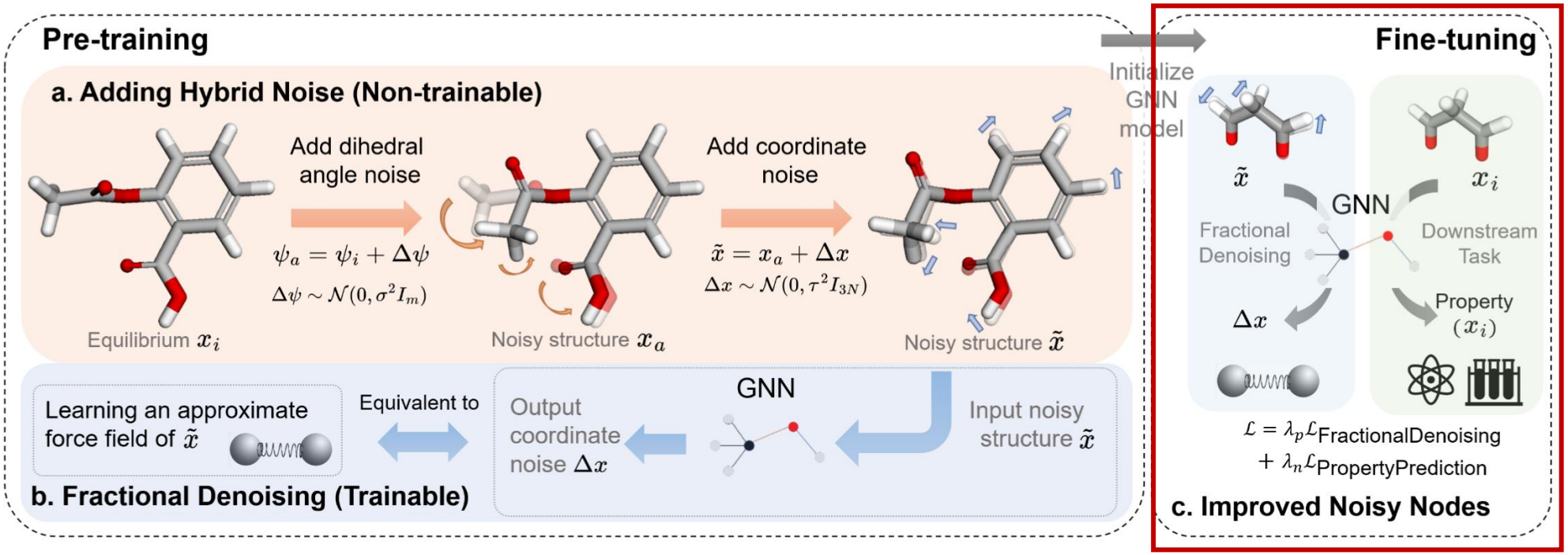


Fractional Denoising Method

$$E_{p(\tilde{x}|x_a)p(x_a|x_i)p(x_i)} ||GNN_\theta(\tilde{x}) - (\tilde{x} - x_a)||^2$$
$$\simeq E_{p(\tilde{x})} ||GNN_\theta(\tilde{x}) - \nabla_{\tilde{x}} \log p(\tilde{x})||^2,$$

- Add hybrid noise, but only reconstruct the coordinate part of noise.
- The fractional denoising task is equivalent to learning the anisotropic force field of hybrid noise.

Frad (Fractional Denoising), ICML 2023



Applying Frad to Fine-tuning

- Noisy Node cannot converge on tasks in MD17 dataset, whereas Noisy Nodes have to corrupt the input conformation leading to an erroneous mapping between inputs and property labels.
- Two modifications:
  - Decouple the denoising task and the downstream task to keep the input of the downstream task unperturbed

Frad (Fractional Denoising), ICML 2023

Table 2. Performance (MAE, lower is better) on QM9. The best results are in bold.

Models	μ (D)	α (a_0^3)	ϵ_{HOMO} (meV)	ϵ_{LUMO} (meV)	$\Delta\epsilon$ (meV)	$< R^2 >$ (a_0^2)	ZPVE (meV)	U_0 (meV)	U (meV)	H (meV)	G (meV)	C_v ($\frac{cal}{molK}$)
SchNet	0.033	0.235	41.0	34.0	63.0	0.07	1.70	14.00	19.00	14.00	14.00	0.033
E(n)-GNN	0.029	0.071	29.0	25.0	48.0	0.11	1.55	11.00	12.00	12.00	12.00	0.031
DimeNet++	0.030	0.043	24.6	19.5	32.6	0.33	1.21	6.32	6.28	6.53	7.56	0.023
PaiNN	0.012	0.045	27.6	20.4	45.7	0.07	1.28	5.85	5.83	5.98	7.35	0.024
SphereNet	0.027	0.047	23.6	18.9	32.3	0.29	**1.120**	6.26	7.33	6.40	8.00	0.022
TorchMD-NET	0.011	0.059	20.3	18.6	36.1	**0.033**	1.840	6.15	6.38	6.16	7.62	0.026
Transformer-M	0.037	0.041	17.5	16.2	**27.4**	0.075	1.18	9.37	9.41	9.39	9.63	0.022
SE(3)-DDM	0.015	0.046	23.5	19.5	40.2	0.122	1.31	6.92	6.99	7.09	7.65	0.024
3D-EMGP	0.020	0.057	21.3	18.2	37.1	0.092	1.38	8.60	8.60	8.70	9.30	0.026
DP-TorchMD -NET($\tau = 0.04$)	0.012	0.0517	17.7	14.3	31.8	0.4496	1.71	6.57	6.11	6.45	6.91	**0.020**
Frad ($\sigma = 2, \tau = 0.04$)	**0.010**	**0.0374**	**15.3**	**13.7**	27.8	0.3419	1.418	**5.33**	**5.62**	**5.55**	**6.19**	**0.020**

SliDe (Sliced Denoising), ICLR 2024

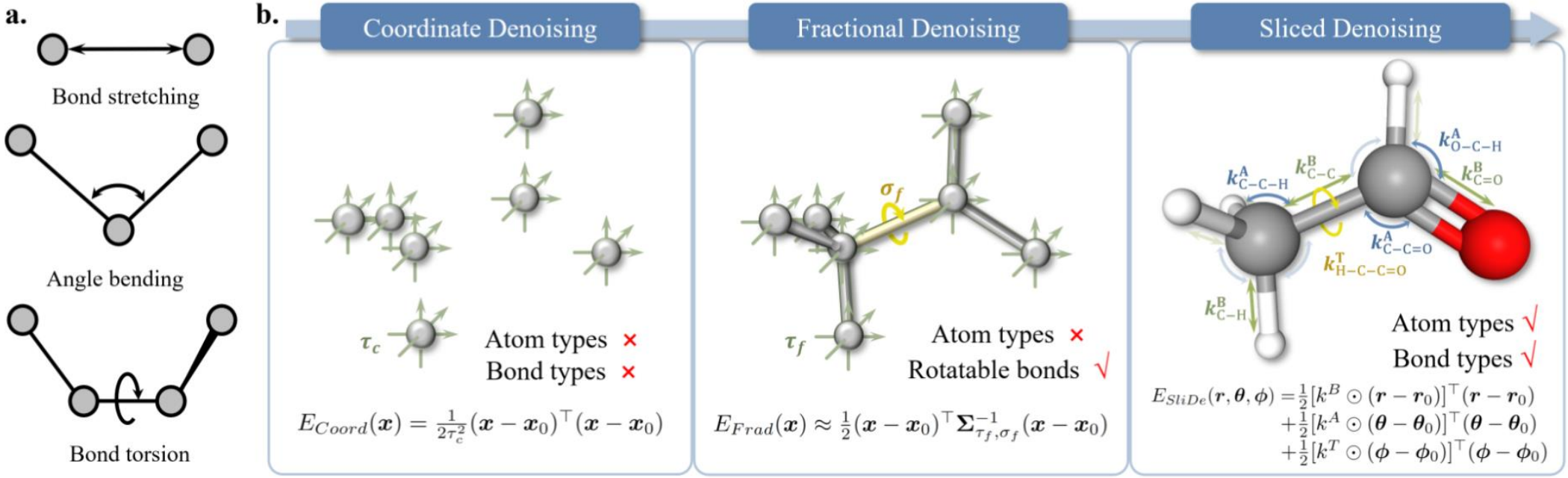
The approximate force field largely deviates from the true force field, due to inappropriate assumptions such as assuming a molecular force field is isotropic in coordinate denoising or treating certain parts being isotropically in Fractional denoising.

Table 11: Performance (MAE ↓) on MD17 force prediction (kcal/mol/ Å). The models are pre-trained on a subset of PCQM4Mv2 dataset.

	Train from Scratch	Coord pre-training	Frad pre-training	DFT label supervised pre-training
Aspirin (Force)	0.253	0.250	0.248	0.236

The energy function is a pivotal factor in determining the quality of representation learning in denoising methods

SliDe (Sliced Denoising), ICLR 2024

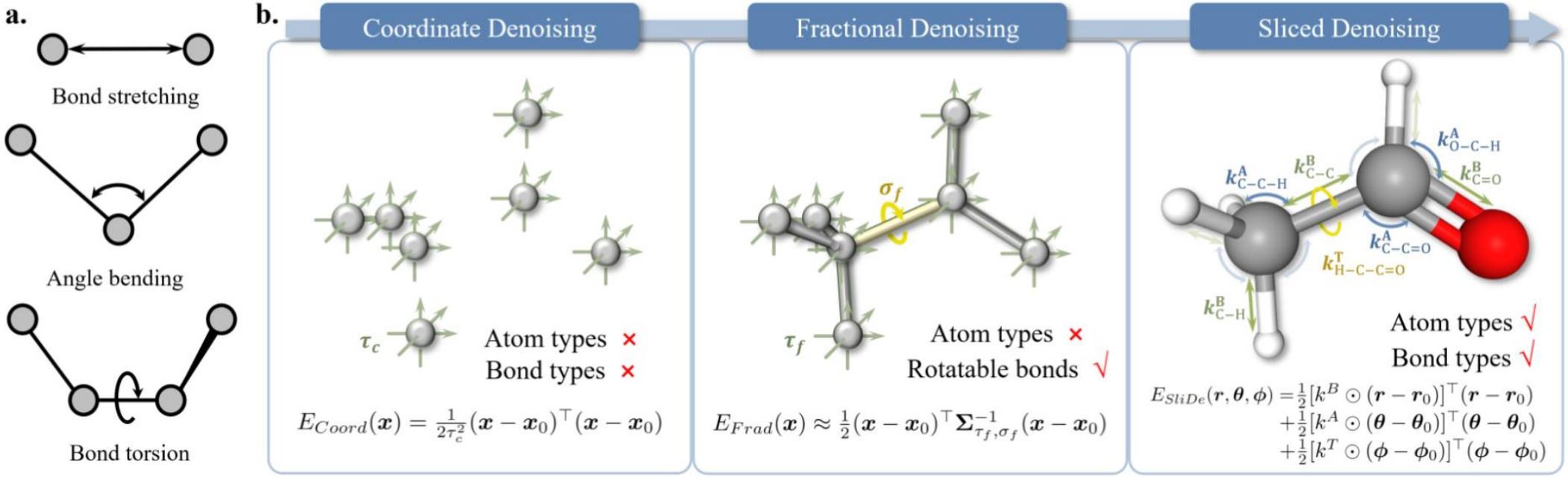


- **The energy function is a pivotal factor** in determining the quality of representation learning in denoising methods

$$\begin{aligned}\mathcal{L}_{Coord}(\mathcal{M}) &= E_{p(\boldsymbol{x}|\boldsymbol{x}_0)p(\boldsymbol{x}_0)}||GNN_\theta(\boldsymbol{x}) - (\boldsymbol{x}-\boldsymbol{x}_0)||^2 \\ &\simeq E_{p(\boldsymbol{x})}||GNN_\theta(\boldsymbol{x}) - (-\nabla_{\boldsymbol{x}}E_{Coord}(\boldsymbol{x}))||^2,\end{aligned}$$

$$\begin{aligned}\mathcal{L}_{Frad}(\mathcal{M}) &= E_{p(\boldsymbol{x}|\boldsymbol{x}_a)p(\boldsymbol{x}_a|\boldsymbol{x}_0)p(\boldsymbol{x}_0)}||GNN_\theta(\boldsymbol{x}) - (\boldsymbol{x}-\boldsymbol{x}_a)||^2 \\ &\simeq E_{p(\boldsymbol{x})}||GNN_\theta(\boldsymbol{x}) - (-\nabla_{\boldsymbol{x}}E_{Frad}(\boldsymbol{x}))||^2.\end{aligned}$$

SliDe (Sliced Denoising), ICLR 2024



- Sliced denoising is based on the **classical mechanical intramolecular potential theory**.
  - Energy function:

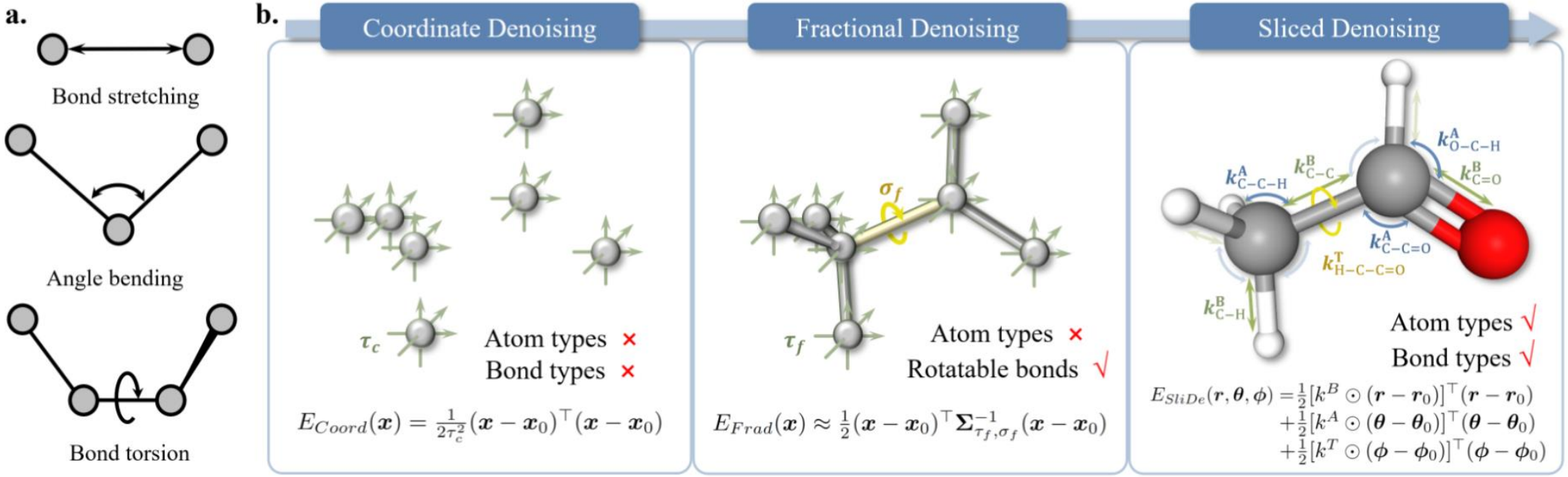
$$E(\boldsymbol{r},\boldsymbol{\theta},\boldsymbol{\phi}) = \frac{1}{2}\sum_{i\in\mathbb{B}} k_i^B(r_i - r_{i,0})^2 + \frac{1}{2}\sum_{i\in\mathbb{A}} k_i^A(\theta_i - \theta_{i,0})^2 + \sum_{i\in\mathbb{T}} k_i^T(1 - cos(\omega_i(\phi_i - \phi_{i,0}))) + E_{elec} + E_{vdW}$$

    - Approximate it as a quadratic form (Taylor expansion).

$$1 - cos(\omega_i(\phi_i - \phi_{i,0})) = 1 - [1 - \frac{1}{2}(\omega_i(\phi_i - \phi_{i,0}))^2 + o((\phi_i - \phi_{i,0})^2)] \approx \frac{1}{2}\omega_i^2(\phi_i - \phi_{i,0})^2$$

    - Drop the last two terms.

SliDe (Sliced Denoising), ICLR 2024



- Sliced denoising is based on the **classical mechanical intramolecular potential theory**.
  - Approximated quadratic form of Energy function:

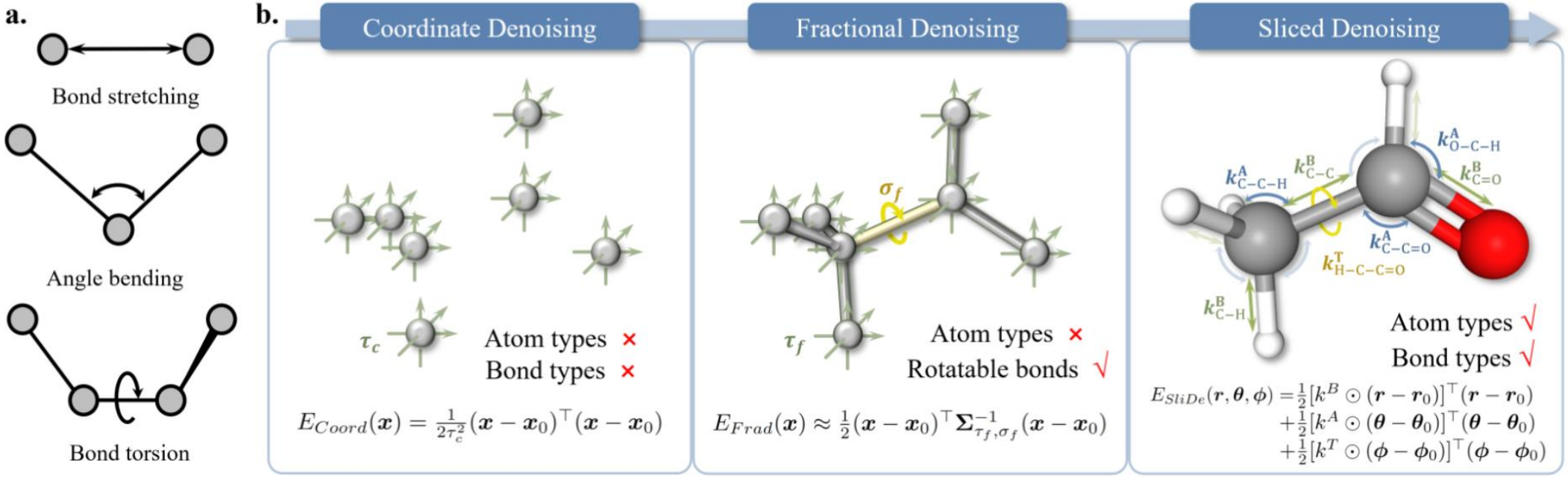
$$E_{BAT}(\boldsymbol{r},\boldsymbol{\theta},\boldsymbol{\phi}) = \frac{1}{2}\sum_{i\in\mathbb{B}} k_i^B(r_i - r_{i,0})^2 + \frac{1}{2}\sum_{i\in\mathbb{A}} k_i^A(\theta_i - \theta_{i,0})^2 + \frac{1}{2}\sum_{i\in\mathbb{T}} k_i^T\omega_i^2(\phi_i - \phi_{i,0})^2$$

  - The conformation distribution of a molecule follows the Boltzmann distribution:

$$p(\boldsymbol{r},\boldsymbol{\theta},\boldsymbol{\phi}) = \frac{1}{Z}exp(-E_{BAT}(\boldsymbol{r},\boldsymbol{\theta},\boldsymbol{\phi})) \tag{7}$$

$$= \prod_{i\in\mathbb{B}}\frac{1}{Z_i^B}exp(-k_i^B\frac{(r_i-r_{i,0})^2}{2})\prod_{i\in\mathbb{A}}\frac{1}{Z_i^A}exp(-k_i^A\frac{(\theta_i-\theta_{i,0})^2}{2})\prod_{i\in\mathbb{T}}\frac{1}{Z_i^T}exp(-k_i^T\omega_i^2\frac{(\phi_i-\phi_{i,0})^2}{2}),$$

SliDe (Sliced Denoising), ICLR 2024



$$E_{Coord}(\boldsymbol{x}) = \frac{1}{2\tau_c^2}(\boldsymbol{x}-\boldsymbol{x}_0)^\top(\boldsymbol{x}-\boldsymbol{x}_0)$$

$$E_{Frad}(\boldsymbol{x}) \approx \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^\top \boldsymbol{\Sigma}^{-1}_{\tau_f,\sigma_f}(\boldsymbol{x}-\boldsymbol{x}_0)$$

$$\begin{aligned} E_{SliDe}(\boldsymbol{r},\boldsymbol{\theta},\boldsymbol{\phi}) = &\frac{1}{2}[k^B \odot (\boldsymbol{r}-\boldsymbol{r}_0)]^\top(\boldsymbol{r}-\boldsymbol{r}_0) \\ &+\frac{1}{2}[k^A \odot (\boldsymbol{\theta}-\boldsymbol{\theta}_0)]^\top(\boldsymbol{\theta}-\boldsymbol{\theta}_0) \\ &+\frac{1}{2}[k^T \odot (\boldsymbol{\phi}-\boldsymbol{\phi}_0)]^\top(\boldsymbol{\phi}-\boldsymbol{\phi}_0) \end{aligned}$$

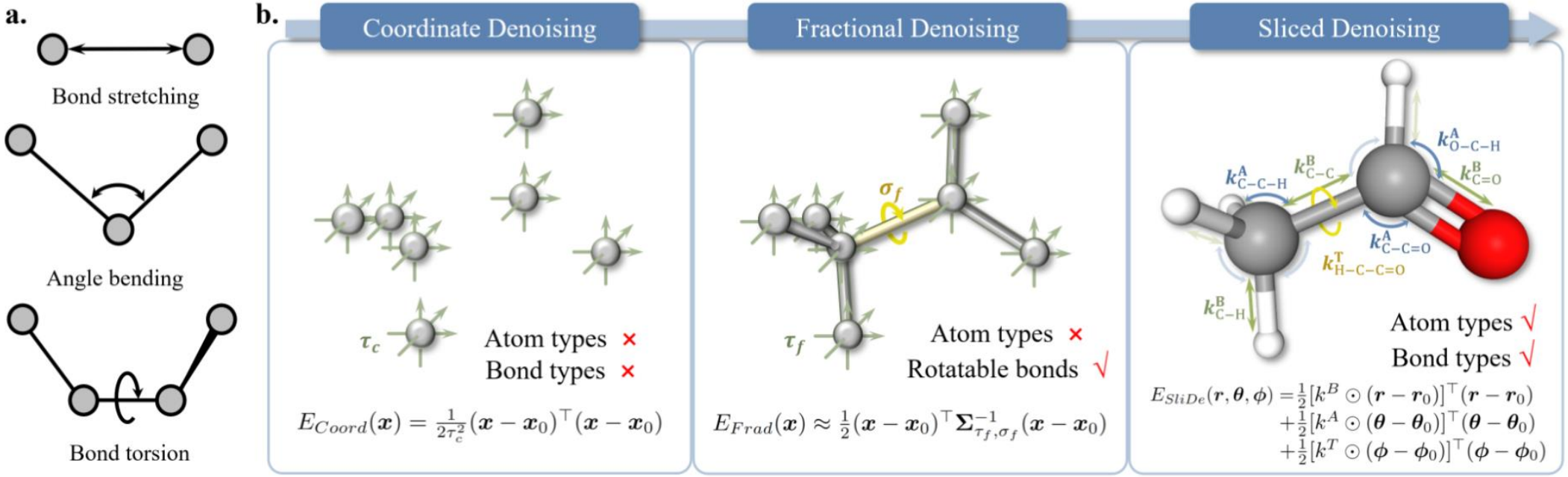
- Sliced denoising is based on the **classical mechanical intramolecular potential theory**.
  - BAT (bond length, angle and torsion angle) noise:

$$\boldsymbol{r} \sim \mathcal{N}(\boldsymbol{r}_0, diag(\frac{1}{\boldsymbol{k}^B})), \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\theta}_0, diag(\frac{1}{\boldsymbol{k}^A})), \boldsymbol{\phi} \sim \mathcal{N}(\boldsymbol{\phi}_0, diag(\frac{1}{\boldsymbol{k}^T \odot \boldsymbol{\omega}^2}))$$

  - However, we need a energy function wrt. Cartesian coordinates to learn the force field:

$$E_{p(\boldsymbol{x}|\boldsymbol{x}_0)}||GNN_\theta(\boldsymbol{x}) - \nabla_{\boldsymbol{x}} E_{BAT}(\boldsymbol{d}(\boldsymbol{x}))||^2$$

SliDe (Sliced Denoising), ICLR 2024



- Sliced denoising is based on the **classical mechanical intramolecular potential theory**.
  - Define a transformation from Cartesian coordinates to relative coordinates:

$$f^{\mathcal{M}} : \mathbb{R}^{3N} \longrightarrow (\mathbb{R}_{\geq 0})^{m_1} \times ([0, 2\pi))^{m_2} \times ([0, 2\pi))^{m_3}$$
$$\boldsymbol{x} \longmapsto \boldsymbol{d} = (\boldsymbol{r}, \boldsymbol{\theta}, \boldsymbol{\phi}),$$

  - Then the force field can be expressed by

$$\nabla_{\boldsymbol{x}} E_{BAT}(f(\boldsymbol{x}))^\top = \nabla_{\boldsymbol{d}} E_{BAT}(\boldsymbol{d})^\top \cdot J(\boldsymbol{x}),$$

where $J(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial f_1^{\mathcal{M}}(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_1^{\mathcal{M}}(\boldsymbol{x})}{\partial x_{3N}} \\ \vdots & & \vdots \\ \frac{\partial f_M^{\mathcal{M}}(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_M^{\mathcal{M}}(\boldsymbol{x})}{\partial x_{3N}} \end{pmatrix} \in \mathbb{R}^{M\times 3N}$ is the Jacobian matrix.

- The target can be written as $E_{p(\boldsymbol{x}|\boldsymbol{x}_0)}||GNN_\theta(\boldsymbol{x}) - \nabla_{\boldsymbol{d}} E_{BAT}(\boldsymbol{d})^\top \cdot J(\boldsymbol{x})||^2.$

SliDe (Sliced Denoising), ICLR 2024

- Sliced denoising is based on the **classical mechanical intramolecular potential theory**.
  - Denoising target:

$$E_{p(\boldsymbol{x}|\boldsymbol{x}_0)}||GNN_\theta(\boldsymbol{x}) - \nabla_{\boldsymbol{d}} E_{BAT}(\boldsymbol{d})^\top \cdot J(\boldsymbol{x})||^2.$$

  - Introduce two computational techniques to achieve unbiased estimation:

Lemma 3.2 (Random Slicing). $\forall \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{v} \in \mathbb{R}^{3N}$, $\sigma > 0$, $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 I_{3N})$, *then*

$$||\boldsymbol{a} - \boldsymbol{b}||^2 = \frac{1}{\sigma^2} E_{\boldsymbol{v}}[(\boldsymbol{a} - \boldsymbol{b})^\top \cdot \boldsymbol{v}]^2.$$

Lemma 3.3 (Differential of Coordinate Transformation Function). $\forall \boldsymbol{x}, \boldsymbol{v} \in \mathbb{R}^{3N}$ *are a molecular conformation and Cartesian coordinate noise respectively, then*

$$J(\boldsymbol{x}) \cdot \boldsymbol{v} = f^{\mathcal{M}}(\boldsymbol{x} + \boldsymbol{v}) - f^{\mathcal{M}}(\boldsymbol{x}) + \alpha(\boldsymbol{x}; \boldsymbol{v}), \tag{16}$$

where $\alpha(\boldsymbol{x}; \boldsymbol{v}) = o(\boldsymbol{v})$ *as* $||\boldsymbol{v}|| \to 0$.

  - Ultimate loss function:

$$\mathcal{L}_{SliDe}(\mathcal{M}) = E_{p(\boldsymbol{x}|\boldsymbol{x}_0)} \frac{1}{N_v} \sum_{i=1}^{N_v} \left[GNN_\theta(\boldsymbol{x})^\top \cdot \boldsymbol{v}_i - \frac{1}{\sigma} \nabla_{\boldsymbol{d}} E_{BAT}(\boldsymbol{d})^\top \cdot \left(f^{\mathcal{M}}(\boldsymbol{x} + \sigma \boldsymbol{v}_i) - f^{\mathcal{M}}(\boldsymbol{x}) \right) \right]^2$$

$$\mathcal{L}_{SliDe}^{total} = \frac{1}{|\mathbb{M}|} \sum_{\mathcal{M} \in \mathbb{M}} \mathcal{L}_{SliDe}(\mathcal{M}).$$

SliDe (Sliced Denoising), ICLR 2024

Evaluation on physical consistency

Table 1: Correlation coefficient between the learned force field and the ground-truth force field of the three methods. The standard deviation is shown in parentheses. The top results are in bold.

Denoising method	Coord	Frad	SliDe
Correlation coefficient	0.616(0.047)	0.631 (0.046)	**0.895** (0.071)

Evaluation on downstream tasks

Table 2: Performance (MAE ↓) on QM9. The best results are in bold.

	μ (D)	α (a_0^3)	homo (meV)	lumo (meV)	gap (meV)	R^2 (a_0^2)	ZPVE (meV)	U_0 (meV)	U (meV)	H (meV)	G (meV)	C_v ($\frac{cal}{mol \cdot K}$)
SchNet	0.033	0.235	41.0	34.0	63.0	0.07	1.70	14.00	19.00	14.00	14.00	0.033
E(n)-GNN	0.029	0.071	29.0	25.0	48.0	0.11	1.55	11.00	12.00	12.00	12.00	0.031
DimeNet++	0.030	0.044	24.6	19.5	32.6	0.33	1.21	6.32	6.28	6.53	7.56	0.023
PaiNN	0.012	0.045	27.6	20.4	45.7	0.07	1.28	5.85	5.83	5.98	7.35	0.024
SphereNet	0.025	0.045	22.8	18.9	31.1	0.27	**1.120**	6.26	6.36	6.33	7.78	0.022
ET	0.011	0.059	20.3	17.5	36.1	**0.033**	1.840	6.15	6.38	6.16	7.62	0.026
TM	0.037	0.041	17.5	16.2	27.4	0.075	1.18	9.37	9.41	9.39	9.63	0.022
SE(3)-DDM	0.015	0.046	23.5	19.5	40.2	0.122	1.31	6.92	6.99	7.09	7.65	0.024
3D-EMGP	0.020	0.057	21.3	18.2	37.1	0.092	1.38	8.60	8.60	8.70	9.30	0.026
Coord	0.012	0.0517	17.7	14.3	31.8	0.4496	1.71	6.57	6.11	6.45	6.91	0.020
Frad	0.010	0.0374	15.3	13.7	27.8	0.3419	1.418	5.33	5.62	5.55	6.19	0.020
SliDe	**0.0087**	**0.0366**	**13.6**	**12.3**	**26.2**	0.3405	1.521	**4.28**	**4.29**	**4.26**	**5.37**	**0.019**